



(11) Publication number : **0 607 054 A2**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number : **94300267.5**

(22) Date of filing : **14.01.94**

(51) Int. Cl.⁵ : **C12N 15/10, C12N 15/12, C12N 15/62, C12N 15/85, C07K 13/00, C07K 15/28, C12Q 1/68, A61K 37/02, A61K 39/395**

(30) Priority : **14.01.93 JP 22098/93**

(43) Date of publication of application :
20.07.94 Bulletin 94/29

(84) Designated Contracting States :
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(71) Applicant : **Honjo, Tasuku**
Kanyu-chi,
Oiwake-cho
Kita-Shirakawa, Sakyo-ku, Kyoto (JP)

(71) Applicant : **ONO PHARMACEUTICAL CO., LTD.**
1-5, Doshomachi 2-chome
Chuo-ku Osaka 541 (JP)

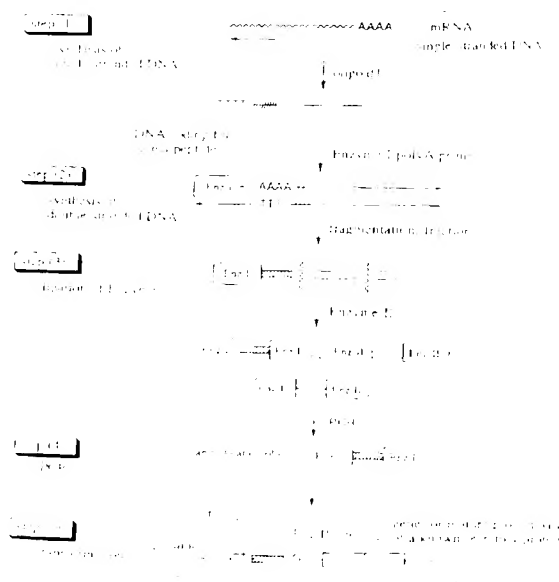
(72) Inventor : **Honjo, Tasuku Kanyu-chi**
Oiwake-cho
Kita-Shirakawa
Sakyo-ku Kyoto (JP)
Inventor : **Tashiro, Kei**
93 Koyama Higashi
Ohno-cho
Kita-ku
Kyoto-shi Kyoto (JP)
Inventor : **Tade, Hideaki c/o Minase Research**
Institute
Ono Pharmaceuticals Co.Ltd.
3-1-1, Sakurai
Shimamoto-cho Mishima-gun Osaka (JP)

(74) Representative : **Bentham, Stephen et al**
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

(54) **Novel process for constructing a cDNA library and a novel polypeptide and DNA coding the same.**

(57) A process for constructing a cDNA library which has a selectivity for signal peptides, that makes it possible to efficiently identify unknown and useful polypeptide comprising a signal peptide. A novel polypeptide consisting of 89 amino acids (including a signal peptide) produced by a stroma cell line, which is useful as an agent for preventing or treating, for example, anemia, leukopenia or infections and the like, and DNAs coding for said polypeptide, have been identified using the process of the invention.

Figure 1



EP 0 607 054 A2

This invention relates to a process for constructing a cDNA library, and a novel polypeptide and a DNA coding for the same. More particularly, it relates to an efficient process for constructing a cDNA library having a high selectivity for signal peptides, and a novel polypeptide produced by a specific stroma cell line and a DNA coding for said polypeptide.

In order to obtain specific polypeptides (for example, proliferation and/or differentiation factors) or a DNA coding for the same, there have been generally employed methods comprising confirming the target biological activity in a tissue or a cell culture medium and then cloning of a gene through the isolation and purification of a polypeptide and further methods comprising expression-cloning of a gene with the guidance of the biological activity.

However, it is frequently observed that a gene, which has been cloned with the guidance of a certain activity, codes for a known polypeptide since many physiologically active polypeptides occurring in vivo have various biological activities. Further, most intravital physiologically active polypeptides are secreted only in a trace amount, which makes the isolation and purification thereof and the confirmation of biological activity difficult.

Recent rapid developments in techniques for constructing cDNAs and sequencing techniques have made it possible to quickly sequence a large amount of cDNAs. By utilising these techniques, a process, which comprises constructing cDNA libraries from various cells and tissues, cloning cDNAs at random, identifying the nucleotide sequences thereof, expressing the corresponding polypeptide and then analyzing its physiological functions, is now in use. Although this process is advantageous in that a gene can be cloned and information on its nucleotide sequence can be obtained without any biochemical or genetic analysis, the target gene can often only be identified by chance.

The present inventors have studied the cloning of genes for proliferation and differentiation factors in hematopoietic and immune systems. They have paid attention to the fact that most secretory proteins such as proliferation and/or differentiation factors (for example various cytokines) and membrane proteins such as receptors thereof (hereinafter these proteins will be referred to generally as secretory proteins and the like) have sequences called signal peptides in the N-termini. Extensive studies have been conducted to provide a process for efficiently and selectively cloning genes coding for signal peptides. As a result, a process has now been devised whereby an N-terminal fragment can be efficiently amplified and the existence of a signal peptide can be easily examined.

In accordance with the present invention, cDNAs with a high probability of containing a signal peptide are ligated at both their ends to linkers containing restriction enzyme sites which are different from each other. These fragments alone are rapidly produced in a large amount by the polymerase chain reaction (PCR) method so as to elevate the content of the fragments with a high probability of containing a signal peptide. Next, the above-mentioned fragment is introduced into an expression vector containing DNA coding for a known secretory protein or the like but lacking DNA encoding the corresponding signal peptide. Many secretory proteins and the like are secreted or expressed on cell membrane even if the signal peptide has been substituted by a signal peptide of another protein. Therefore, if the known secretory protein or the like is expressed on cell membrane or outside the cells, this confirms that a cDNA fragment corresponding to a signal peptide has been introduced into the expression vector. Thus the inventors have devised a convenient method for detection of the insertion of cDNA encoding a signal peptide.

The polymerase chain reaction used in the present invention is known as a method for amplifying specific DNA fragments in a large amount. It is also known that many secretory proteins and the like can be expressed even if the signal peptide thereof is substituted by that of another secretory protein or the like. However, to the applicants' knowledge, there has been no suggestion that these techniques be used as the inventors have done to provide a process for selectively cloning a signal peptide.

It is known that hematopoietic cells secrete various proliferation and/or differentiation factors exemplified by interleukin.

The present invention further relates to a novel polypeptide obtained from hematopoietic cells and DNA coding for the same.

The inventors have sought a novel factor (polypeptide) produced by hematopoietic cells using the process which is the first subject of the present invention. As a result, a novel polypeptide and DNA coding for the same have been identified.

Computer searches to identify polypeptides having sequences identical or highly homologous with that of the polypeptide of the present invention and the DNAs coding for the same have not identified any such sequences. Thus, to the best of the applicants' knowledge, the polypeptide of the present invention and the DNA coding for the same are novel. Further, homology analysis has revealed that the polypeptide of the present invention is a member of the chemokine family as it has a pattern of Cys-X-Cys (X is optional amino acid).

Accordingly, the present invention provides a process for constructing a cDNA library comprising:

sites may be used as the restriction enzyme (enzyme I) site to be ligated to a poly A oligomer serving as a primer and the restriction enzyme (enzyme II) site to be used in the next step (3), so long as they differ from each other. It is preferable to use EcoRI and SacI, as enzyme I and enzyme II respectively.

In the step (3), the double-stranded DNA is fragmented for example by ultrasonication so as to give an average cDNA length of 300 bp and the obtained fragments are fractionated into cDNAs of 200 to 500 bp by agarose gel electrophoresis (AGE). After blunting with T4DNA polymerase, enzyme II is ligated and the cDNAs are fractionated into DNAs of 200 to 500 bp again by agarose gel electrophoresis. As described above, any enzyme may be used as the enzyme II, so long as it differs from the enzyme I. The procedure in this step increases the likelihood that a cDNA fragment containing a signal peptide exists in the part located between the enzymes I and II.

In the step (4), PCR is carried out in order to further elevate the possibility that a cDNA fragment containing a signal peptide exists in the part located between the enzymes I and II. PCR is a well known technique and automated devices therefore are commercially available. It is sufficient to amplify 25 to 30 times. The cDNA thus amplified is digested with enzyme I and enzyme II and electrophoresed on an agarose gel to thereby fractionate into cDNAs of 200 to 500 bp.

In the step (5), a gene for a known secretory protein or the like with the deletion of signal peptide, which is called a reporter gene, and a cDNA fragment obtained in the above (4) are integrated into an eucaryotic cell expression plasmid vector in such a manner that the cDNA fragment is located upstream of the reporter gene. This is followed by transformation of the vector into a host.

Various known eucaryotic cell expression plasmid vectors for example, pcDL-SR α and pcEV-4 which are capable of acting in *Escherichia coli*, are usable in the present invention.

As the reporter gene, genes for mature protein parts of soluble secretory proteins and membrane proteins of any type are usable. The expression of these reporter genes may be confirmed by known methods such as antibody methods. Human IL-2 receptor α gene is especially suitable as a reporter gene.

A number of *E. coli* strains are known as hosts for transformation and any of these strains is usable. It is preferable to use DH5 competent cells [described in Gene, 96, 23 (1990)] therefor. Transformants may be incubated in a conventional manner and thus the cDNA library of the present invention can be obtained.

In the process for constructing a cDNA library according to the present invention, there is a high possibility that gene fragments coding for signal peptides are contained in the library. However, not every clone contains said fragment. Further, not all of the gene fragments code for unknown (novel) signal peptides. It is therefore necessary to screen a gene fragment coding for an unknown signal peptide from said library.

Namely, the cDNA library is divided into small pools of an appropriate size and integrated into an expression system. Examples of the expression system for producing a polypeptide include mammalian cells (for example, monkey COS-7 cells, Chinese hamster CHO cells, mouse L cells, etc.). Transfection may be performed in accordance with well known methods such as the DEAE-dextran method. After the completion of the incubation, the expression of the reporter gene is examined. It is known that a reporter gene would be expressed even though the signal peptide is the one characteristic to another secretory protein. That is to say, the fact that the reporter gene has been expressed indicates that a signal peptide of some secretory protein has been integrated into the library. Positive pools are further divided into smaller ones and the expression and the judgement are repeated until a single clone is obtained. The expression of the reporter gene can be judged by, for example, fluorescence-labeled antibody assay, enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA), depending on kinds of the employed reporter gene.

Next, the nucleotide sequence of the isolated positive clone is determined. In the case of a cDNA which is proved to code for an unknown protein, the clone of the full length is isolated with the use of the cDNA as a probe and the full nucleotide sequence can be thus identified. All of these operations are carried out by methods which are well known by those skilled in the art. For example, the nucleotide sequence may be identified by the Maxam-Gilbert method or the dideoxy terminator method. On the other hand, the full length may be sequenced in accordance with a method described in Molecular Cloning [Sambrook, J., Fritsch, E.F. and Maniatis, T., published by Cold Spring Harbor Laboratory Press in 1989].

The present invention further relates to a novel polypeptide which has been identified using the process of the present invention to construct a cDNA library and a DNA coding for the same. In particular, it relates to:

- (1) a polypeptide consisting of an amino acid sequence represented by SEQ ID No. 1;
- (2) a DNA coding for the polypeptide described in the above (1);
- (3) a DNA having a nucleotide sequence represented by SEQ ID No. 2; and
- (4) a DNA having a nucleotide sequence represented by SEQ ID No. 3.

A polypeptide of Seq. ID No. 1 in substantially purified form will generally comprise the polypeptide in a preparation in which more than 90%, eg. 95%, 98% or 99% of the polypeptide in the preparation is that of the Seq. ID No. 1.

A polypeptide homologue of the Seq. ID No. 1 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the polypeptide of Seq. ID No. 1 over a region of at least 20, preferably at least 30, for instance 40, 60 or 100 more contiguous amino acids. Such polypeptide homologues will be referred to below as a polypeptide according to the invention.

Generally, fragments of Seq. ID No. 1 or its homologues will be at least 10, preferably at least 15, for example 20, 25, 30, 40, 50 or 60 amino acids in length, and are also encompassed by the term "a polypeptide according to the invention" as used herein. particular fragments of the polypeptides of the invention are fragments of which include amino acid residues numbered 1-70 in Seq ID No. 4 or a homologue thereof.

A DNA capable of selectively hybridizing to the DNA of Seq. ID No. 2 or 3 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the DNA of Seq. ID No. 2 or 3 over a region of at least 20, preferably at least 30, for instance 40, 60 or 100 or more contiguous nucleotides. Such DNA will be encompassed by the term "DNA according to the invention".

Particular DNA capable of selectively hybridising to the DNA of SEQ ID No. 2 or 3 is the nucleotide residues numbered 139-348 in SEQ ID No. 4 or a fragment thereof.

DNA according to the invention may be used to produce a primer, eg. a PCR primer, a probe eg. labelled by conventional means using radioactive or non-radioactive labels, or the DNA may be cloned into a vector. Such primers, probes and other fragments of the DNA of Seq. ID No. 2 or 3 will be at least 15, preferably at least 20, for example 25, 30 or 40 nucleotides in length, and are also encompassed by the term "DNA according to the invention" as used herein.

DNA according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art.

A further embodiment of the invention provides replication and expression vectors comprising DNA according to the invention. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said DNA and optionally a regulator of the promoter. The vector may contain one or more selectable marker genes, for example an ampicillin resistance gene. The vector may be used *in vitro*, for example for the production of RNA corresponding to the DNA, or used to transform a host cell.

A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of DNA according to the invention, including the DNA Seq. ID No. 2 or 3 or the open reading frame thereof. The cells will be chosen to be compatible with the vector and may for example be bacterial, yeast, insect or mammalian.

DNA according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA (DNA). Antisense RNA (DNA) may also be produced by synthetic means. Such antisense RNA (DNA) may be used in a method of controlling the levels of a polypeptide of the invention in a cell.

A further embodiment of the invention provides a method of producing a polypeptide which comprises culturing host cells of the present invention under conditions effective to express a polypeptide of the invention. Preferably, in addition, such a method is carried out under conditions in which the polypeptide of the invention is expressed and then secreted from the host cells.

The invention also provides monoclonal or polyclonal antibodies to a polypeptide according to the invention. The invention further provides a process for the production of monoclonal or polyclonal antibodies to the polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using a polypeptide of the invention or a fragment thereof, as an immunogen. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention and recovering immune serum.

The present invention also provides pharmaceutical compositions containing a polypeptide of the invention, or an antibody thereof, in association with a pharmaceutically acceptable diluent or carrier.

The invention also provides a polypeptide according to the invention or an antibody for use in a method of therapy or diagnosis on a human or animal body.

The polypeptides of the present invention include not only those having the amino acid sequence represented by the SEQ ID No. 1 but also those with partial deletion thereof (for example, a polypeptide consisting of the mature protein part alone, or consisting of a part of the mature protein essentially required for the expression of the biological activity), those with partial replacement by other amino acid(s) (for example, a polypeptide some of amino acids are replaced by those having similar properties) and those with partial addition or insertion of amino acid(s).

It is well known that there are up to six different codons which may code for a single amino acid (for example, one type of codon for Met while six types of codon for Leu). Accordingly, the nucleotide sequence of the DNA can be changed without altering the amino acid sequence of the polypeptide.

The DNA as specified in (2) includes all nucleotide sequences coding for the polypeptide represented by Seq. ID No. 1. Changes in the nucleotide sequence sometimes bring about an increase in the polypeptide productivity.

The DNA as specified in (3) is an embodiment of the DNA as specified in (2) and represents a natural sequence.

The DNA as specified in (4) represents a sequence wherein a natural non-translational region is added to the DNA as specified in (3).

A signal peptide is a highly hydrophobic region located immediately downstream of the translation initiation amino acid Met. It is assumed that the signal peptide in the polypeptide of the present invention resides in a region ranging from Met at the 1-position to Ser at the 19-position in the amino acid sequence represented by Seq. ID No. 1. The region essentially responsible for the expression of the biological activity corresponds to the part of the amino acid sequence of the Seq. ID No. 1 lacking of the signal peptide, i.e., the mature protein part. Thus the signal peptide never relates to the activity.

The DNA having the nucleotide sequence represented by Seq. ID No. 3 can be prepared in accordance with the process described as the first subject of the present invention.

Once the nucleotide sequences represented by Seq. ID No. 2 and No. 3 are determined, the DNA of the present invention can be chemically synthesized. Alternatively, the DNA of the present invention can be obtained by chemically synthesizing fragments of said nucleotide sequence and hybridizing with the use of the fragments as a probe. Further, the target DNA can be produced in a desired amount by introducing a vector DNA containing said DNA into an appropriate host and then incubating the host.

Examples of methods for obtaining the polypeptide of the present invention include:

- (1) isolation and purification from vital tissues or cultured cells;
- (2) chemical synthesis of peptides; and
- (3) production with the use of gene recombination techniques. From an industrial viewpoint, the method described in (3) is preferable.

Examples of the expression system (host-vector system) for producing the polypeptide by using gene recombination techniques include those of bacteria, yeasts, insect cells and mammalian cells.

In order to express in *E. coli*, for example, an initiator codon (ATG) is added to the 5'-end of the DNA coding for the mature protein region. The DNA thus obtained is then ligated to the downstream of an appropriate promoter (for example, trp promoter, lac promoter, λ p_L promoter, T7 promoter, etc.) and inserted into a vector capable of functioning in *E. coli* (for example, pBR322, pUC18, pUC19, etc.), thus constructing an expression vector. Next, an *E. coli* strain (for example, *E. coli* DH1, *E. coli* JM109, *E. coli* HB101, etc.) transformed with this expression vector is incubated in an appropriate medium. Thus the target polypeptide can be obtained from the incubated cells. Alternately, a bacterial signal peptide (for example, a signal peptide of pelB) may be used and thus the polypeptide can be secreted into the periplasm. Furthermore, a fusion protein together with other polypeptide can be produced.

Expression in mammalian cells can be effected, for example, in the following manner. Namely, a DNA coding for the nucleotide sequence represented by Seq. ID No. 3 is inserted into the downstream of an appropriate promoter (for example, SV40 promoter, LTR promoter, metallothionein promoter, etc.) in an appropriate vector (for example, retrovirus vector, papilloma virus vector, vaccinia virus vector, SV40-series vector, etc.), thus constructing an expression vector. Next, appropriate mammalian cells (for example, monkey COS-7 cells, Chinese hamster CHO cells, mouse L cells, etc.) are transformed with the expression vector obtained above and the transformant is incubated in an appropriate medium. Thus the target polypeptide can be secreted into the culture medium. The polypeptide thus obtained can be isolated and purified by conventional biochemical methods.

By using the process for constructing a cDNA library which is the first subject of the present invention, a DNA coding for a signal peptide of a secretory protein or a membrane protein can be efficiently selected and, in its turn, an unknown and useful protein can be efficiently found out. The novel polypeptide which is the second subject of the present invention is produced and secreted from a stroma cell line. Therefore, the polypeptide has biological activities relating to the survival and proliferation of hematopoietic stem cells and the proliferation and differentiation of B cells and myeloid cells, and chemoattractant activity of neurophil. Accordingly, the polypeptide of the present invention per se is usable as an agent for preventing or treating, for example, anemia or leukopenia, infections, etc.

In addition, the above-mentioned polypeptide existing in vivo can be assayed by using a polyclonal antibody or a monoclonal antibody for said polypeptide, which is applicable to studies on the relationship between said polypeptide and diseases or to the diagnosis of diseases and the like. The polyclonal antibody and the monoclonal antibody can be prepared by a conventional method with the use of said polypeptide or a fragment thereof as an antigen.

The DNA according to the present invention serves as an important and essential template in the production of the polypeptide of the present invention which is expected to be highly useful. Further, the DNA of the present invention is applicable to the diagnosis and treatment of hereditary diseases, i.e., gene therapy, and therapy with ceasing the expression of the polypeptide by using antisense DNA (RNA). Furthermore, a genomic DNA can be isolated by using the DNA of the present invention as a probe. Similarly, a human gene for a related polypeptide being highly homologous with the DNA of the present invention and a gene of an organism other than human for a polypeptide being high homologous with the polypeptide of the present invention can be isolated.

Examples

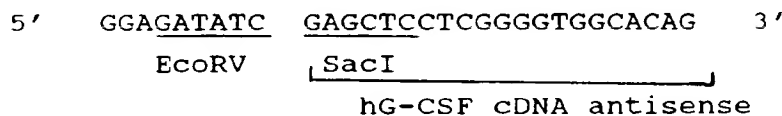
The following Examples and Reference Example are illustrated, but not limit the present invention.

Reference Example 1

Construction and expression of plasmid pcDL-SR α -h-G-CSF-hTac (pSGT) and plasmid pcDL-SR α -hRAR α -hTac (pSRT)

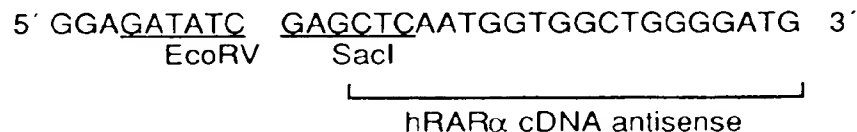
A plasmid, wherein a cDNA coding for a fusion protein of hG-CSF (human granulocyte colony stimulating factor, a typical example of a protein having a signal peptide) or hRAR α (human retinoic acid receptor α , a typical example of a protein having no signal peptide), with hTac (human IL-2 receptor α , used as a reporter gene) was integrated into an eucaryotic cell expression plasmid vector pcDL-SR α having an SR α promoter [described in Mol. Cell. Biol., 8, 466 (1988), provided by Dr. Yutaka Takebe, National Institute of Health], was constructed. After transformation, the expression of the reporter protein on the membrane was examined.

(1) By employing a plasmid pSP72-hG-CSF, wherein hG-CSF cDNA had been integrated into the EcoRI site of a plasmid pSP72 (purchased from Promega), as a template and using an SP6 promoter primer (purchased from Takara Shuzo Co., Ltd.) and an hG-CSF specific primer having an SacI site added thereto,



PCR was performed 25 cycles (at 95 °C for one minute, at 48 °C for two minutes and at 72 °C for two minutes). The amplified DNA fragment was digested with SacI-EcoRI and once subcloned into a plasmid pBlue script SK(+) (pBS). After digesting with SacI-EcoRI again, an EcoRI-SacI fragment of hG-CSF was obtained. On the other hand, a plasmid pBS-hTac, wherein hTac cDNA had been integrated into the HindIII site of pBS, was digested with SacI-KpnI to thereby give an SacI-KpnI fragment of hTac cDNA with the deletion of the signal sequence. These fragments were integrated into the EcoRI-KpnI site of pcDL-SR α with the deletion of stuffer (Fig. 2) to thereby give a plasmid pcDL-SR α -hG-CSF-hTac (pSGT) (Figs. 3, 4 and 5).

Next, by employing a plasmid pGEM3-hRAR α , wherein hRAR α cDNA had been integrated into the EcoRI site of a plasmid pGEM3, as a template and using an SP6 promoter primer and an hRAR α specific primer having an SacI site added thereto,



PCR was performed. Subsequently, the procedure employed in the above-mentioned case of G-CSF was repeated to thereby give a plasmid pcDL-SR α -hRAR α -hTac (pSRT) (Figs. 6, 4 and 5).

(2) The pSGT or pSRT obtained in the above (1) was transfected into COS-7 cells by the DEAE-dextran method [described in detail in Current Protocol in Molecular Biology, chapter 9.2.1.]. After 48 hours, the cells were harvested from a dish and incubated together with mouse anti-Tac igG antibody for 20 minutes on ice. After eliminating free antibodies, the mixture was incubated together with goat anti-mouse igG an-

tibody labeled with fluorescein isothiocyanate (FITC) for 20 minutes on ice. After eliminating free antibodies again, the expression of a fusion protein G-CSF-Tac or RAR α -Tac on the membrane was examined with a fluorescence activated cell sorter (Model FACS Can, manufactured by BECTON DICKINSON, hereinafter referred to simply as FACS). Figs. 7 and 8 show the results of the judgement.

As shown in Fig. 7, G-CSF-Tac was expressed on the membrane as well as Tac. As shown in Fig. 8, on the other hand, RAR α -Tac was not detected on the membrane but remained within the cells. These results indicate that when a cDNA fragment containing a signal peptide is ligated to the upstream of a reporter gene, said reporter protein is expressed on the cell membrane, while when a cDNA fragment having no signal peptide is ligated, the reporter protein is not expressed.

Example 1

Construction of cDNA library having selectivity for signal peptides (Figs. 9 and 10)

Total RNA was extracted from a mouse stroma cell line ST2 [cells supporting the survival and proliferation of hematopoietic stem cells and the proliferation and differentiation of B cells and myeloid cells; refer to EMBO J., 7, 1337 (1988)] by the acid guanidine-phenol-chloroform (AGPC) method [described in detail in "Saibo Kogaku Jikken Protokoru (Protocol in Cellular Engineering Experiments)", published by Shujun-sha, 28 - 31]. Then poly A-RNA was purified by using oligo (dT)-latex (Oligotex-dT30®, marketed from Takara Shuzo Co., Ltd.). By using a random hexamer as a primer, a single-stranded cDNA was synthesized with reverse transcriptase and dT was added to the 3'-end thereof with the use of terminal deoxytransferase. A 17 mer dA ligated to a restriction enzyme site containing EcoRI

5' GATGCGGCCGC CTCTGAG GAATTC (dA)₁₇ 3'
NotI XhoI EcoRI

was annealed and a double-stranded cDNA was synthesized by using the same as a primer. Then the cDNA was fragmented by ultrasonication so as to give an average length of 300 bp and the cDNAs of 200 to 500 bp were fractionated by agarose gel electrophoresis. After blunting the ends with T4DNA polymerase, a linker containing an SacI site

5' GAGGTACAAGCTT GATAIC GAGCTCGCGG 3'
3' CATGTTCTGAA CTATAG CTCGAGCGCC 5'
Hind III EcoRV SacI

[see Nucleic Acids Res., 18, 4293 (1990)] was ligated and cDNAs of 200 to 500 bp were fractionated again by agarose gel electrophoresis. By using a primer (NLC) containing an EcoRI site

NLC

5' GATGCGGCCGCCTCTGAGGAATTC 3'

and another primer (LLHES) containing an SacI site

LLHES

5' GAGGTACAAGCTTGATATCGAGCTCGCGG 3'

PCR was performed 25 cycles (at 94 °C for one minute, at 50 °C for two minutes and at 72 °C for two minutes). The amplified cDNA was digested with SacI and EcoRI and cDNAs of 250 to 500 bp were fractionated by agarose gel electrophoresis. The cDNA was ligated to a plasmid obtained by digesting pSRT (prepared in Reference Example 1) with SacI and EcoRI by using T4 DNA ligase. After transformation of an *E. coli* DH5 α strain, a cDNA library having a selectivity for signal peptides was obtained.

Example 2

Screening and analysis of cDNA coding for signal peptide

About 1,200 colonies in the library obtained in Example 1 were divided into 24 pools (about 50 colonies/pool).
 5 Plasmids of each pools were isolated by the miniprep method and transfected into COS-7 cells by the DEAE-dextran method. After 48 to 72 hours, cell surface-staining for Tac of the transfectant was performed in the same manner as described in Reference Example 1 and 6 positive pools were selected under a fluorescent microscope. Colonies of one pool from among the 6 positive pools were further divided and the same procedure
 10 as described above was repeated until a single clone was obtained. Thus a positive clone (pS-TT3) was obtained. Subsequently, by using two synthetic primers

5' TTTACTTCTAGGCCTGTACG 3'

(20 bases upstream from EcoRI cloning site, for sense)

15 and

5' CCATGGCTTTGAATGTGGCG 3'

(20 bases downstream from SacI cloning site, for antisense)

20 which were specific for the pcDL-SR α -Tac vector, the nucleotide sequence of the TT3 insert was determined. An open reading frame following the Tac cDNA with the deletion of the signal sequence in-frame was searched and converted into the deduced amino acid sequence. After performing a hydrophobicity profile, it was confirmed that a hydrophobic region characteristic to a signal peptide was contained therein (Fig. 11). Further, the homology with data base on DNA and amino acid levels was examined. As a result, it has been found out
 25 that TT3 clone codes for an unknown protein.

Example 3

Screening of cDNA with the full length and determination of nucleotide sequence

30 A cDNA library was constructed by using Super Script[®] Ramda System (marketed from BRL). Next, pS-TT3 was digested with SacI and EcoRI and a TT3 cDNA fragment was prepared by agarose gel electrophoresis. The library was screened by using an oligo-labeled TT3 cDNA fragment as a probe and thus a number of positive clones were obtained. Among these clones, a TT3-1-6 clone showing the longest insert was selected and an Sall-NotI fragment excised from a λ gt22A vector was subcloned into pBS to thereby give a plasmid pBS-
 35 TT316. By using a T7 primer, the nucleotide sequence of 300 bp in the 5'-terminal of TT3-1-6 cDNA was determined. Thus it was confirmed that the sequence identical with TT3 of the probe existed in the most 5'-end of TT3-1-6.

Next, a number of pBS-TT316 variant plasmids lacking of the 5'-end or the 3'-end of TT3-1-6 cDNA were constructed by using an ExoIII/Mung Bean Deletion Kit (manufactured by Stratagene). By using these variant
 40 plasmids, the nucleotide sequence of the full length of the cDNA was determined (sequence No. 3). From the full length cDNA sequence data, an open reading frame was determined and further translated into an amino acid sequence. Thus the sequence represented by Seq. ID No. 1 was obtained. The amino acid sequence at the 30- to 40-positions in the N-terminal of the amino acid sequence thus obtained were compared with known signal peptides. Thus the signal peptide part of this polypeptide was deduced to thereby give the sequence
 45 represented by Seq. ID No. 4.

50

55

SEQUENCE LISTING

5 GENERAL INFORMATION:

APPLICANT:

NAME: Ono Pharmaceutical Co., Ltd.

10 STREET: 1-5 Doshomachi 2-chome

CITY: Osaka

15 COUNTRY: Japan

POSTAL CODE (ZIP): 541

20 TITLE OF THE INVENTION:

Process for constructing cDNA library, and novel polypeptide and

25 DNA coding for the same

NUMBER OF SEQUENCES: 4

30 SEQUENCE NO: 1

LENGTH: 89 amino acids

TYPE: amino acid

35 TOPOLOGY: linear

MOLECULE TYPE: protein

40 SEQUENCE: Seq. ID No. 1

Met Asp Ala Lys Val Val Ala Val Leu Ala Leu Val Leu Ala Ala Leu
 1 5 10 15
 45 Cys Ile Ser Asp Gly Lys Pro Val Ser Leu Ser Tyr Arg Cys Pro Cys
 20 25 30
 Arg Phe Phe Glu Ser His Ile Ala Arg Ala Asn Val Lys His Leu Lys
 35 40 45
 Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln Ile Val Ala Arg Leu Lys
 50 55 60
 Asn Asn Asn Arg Gln Val Cys Ile Asp Pro Lys Leu Lys Trp Ile Gln
 65 70 75 80
 Glu Tyr Leu Glu Lys Ala Leu Asn Lys
 85

55

SEQUENCE NO: 2

LENGTH: 267 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

SEQUENCE: Seq. ID No. 2

```

15 ATGGACGCCA AGGTCGTCGC CGTGCTGGCC CTGGTGCTGG CCGCGCTCTG CATCAGTGAC 60
   GGTAAACCAG TCAGCCTGAG CTACCGATGC CCTGCGCGGT TCTTCGAGAG CCACATCGCC 120
   AGAGCCAACG TCAAGCATCT GAAAATCCTC AACACTCCAA ACTGTGCCCT TCAGATTGTT 180
20 GCACGGCTGA AGAACAACAA CAGACAAGTG TGCATTGACC CGAAATTAAA GTGGATCCAA 240
   GAGTACCTGG AGAAAGCTTT AAACAAG 267

```

SEQUENCE NO: 3

LENGTH: 1797 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

SEQUENCE: Seq. ID No. 3

```

40 GACCACTTTC CCTCTCGGTC CACCTCGGTG TCCTCTTGCT GTCCAGCTCT GCAGCCTCCG 60
   GCGCGCCCTC CCGCCACGC CATGGACGCC AAGGTCGTCG CCGTGCTGGC CCTGGTGCTG 120
   GCGCGCTCT GCATCAGTGA CGGTAAACCA GTCAGCTGA GCTACCGATG CCTTGCCGG 180
45 TTCTTCGAGA GCCACATCGC CAGAGCCAAC GTCAAGCATC TGAAAATCCT CAACACTCCA 240
   AACTGTGCCC TTCAGATTGT TGCACGGCTG AAGAACAACA ACAGACAAGT GTGCATTGAC 300
   CCGAAATTAA AGTGGATCCA AGAGTACCTG GAGAAAGCTT TAAACAAGTA AGCACAACAG 360
50 CCCAAAGGAC TTTCCAGTAG ACCCCCGAGG AAGGCTGACA TCCGTGGGAG ATGCAAGGGC 420
   AGTGGTGGGG AGGAGGGCCT GAACCCTGGG CAGGATGGCC GGCGGGACAG CACTGACTGG 480
   GGTGATGCTA AGGTTTGCCA GCATAAAGAC ACTCCGCCAT AGCATATGGT ACGATATTGC 540
   AGCTTATATT CATCCCTGCG CTCGCCCGTG CACAATGGAG CTTTATAAC TGGGGTTTTT 600
55 CTAAGGAATT GTATTACCT AACCAGTTAG CTTGATCCGC ATTCTCCTCA TCCTCATCTT 660

```

CATTTTAAAA AGCAGTGATT ACTTCAAGGG CTGTATTCAG TTTGCTTTGG AGCTTCTCTT 720
 TGCCCTGGGG CCTCTGGGCA CAGTTATAGA CGGTGGGTTT GCAGGGAGCC CTAGAGAGAA 780
 ACCTTCCACC AGAGCAGAGT CCGAGGAACG CTGCAGGGCT TGCTCTGCAG GGGGGGCTCC 840
 5 TCGACAGATG CTTTGTCTTG AGTCAACACA AGATCCGGCA GAGGGAGGCT CCTTTATCCA 900
 GTTCAGTGCC AGGGTCGGGA AGCTTCCTTT AGAAGTGATC CCTGAAGCTG TGCTCAGAGA 960
 CCCTTTCTTA GCCGTTCTTG CTCTCTGCTT GCCTCCAAAC GCATGCTTCA TCTGACTTCC 1020
 10 GCTTCTCACC TCTGTAGCCT GACGGACCAA TGCTGCAATG GAAGGGAGGA GAGTGATGTG 1080
 GGGTGCCCCC TCCCTCTCTT CCCTTTGCTT TCCTCTCACT TGGGCCCTTT GTGAGATTTT 1140
 TCTTTGGCCT CCTGTAGAAT GGAGCCAGAC CATCCTGGAT AATGTGAGAA CATGCCTAGA 1200
 15 TTTACCCACA AAACACAAGT CTGAGAATTA ATCATAAACG GAAGTTTAAA TGAGGATTTG 1260
 GACCTTGGTA ATTGTCCCTG AGTCCTATAT ATTTCAACAG TGGCTCTATG GGCTCTGATC 1320
 GAATATCAGT GATGAAAATA ATAATAATAA TAATAATAAC GAATAAGCCA GAATCTTGCC 1380
 ATGAAGCCAC AGTGGGGATT CTGGGTTCCA ATCAGAAATG GAGACAAGAT AAAACTTGCA 1440
 20 TACATTCTTA TGATCACAGA CGGCCCTGGT GGTTTTTGGT AACTATTTAC AAGGCATTTT 1500
 TTTACATATA TTTTGTGCA CTTTTTATGT TTCTTTGGAA GACAAATGTA TTTCAGAATA 1560
 TATTTGTAGT CAATTCATAT ATTTGAAGTG GAGCCATAGT AATGCCAGTA GATATCTCTA 1620
 25 TGATCTTGAG CTACTGGCAA CTTGTAAAGA AATATATATG ACATATAAAT GTATTGTAGC 1680
 TTTCCGCTGT CAGCCACGGT GTATTTTTC ACTTGGAATG AAATGTATC AACTGTGACA 1740
 TTATATGCAC TAGCAATAAA ATGCTAATTG TTTCATGCTG TAAAAAAAAA AAAAAAA 1797

30

SEQUENCE NO: 4

LENGTH: 1797 base pairs

35

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

40

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE

45

ORGANISM: Mouse

CELL LINE: ST2

FEATURE

50

NAME/KEY: CDS

LOCATION: 82 .. 351

55

IDENTIFICATION METHOD: P

5

10

15

25

36

35

46

15

50

55

CTTTATCCAG TTCASTGCCA GGGTCGGGAA GCTTCCTTTA GAAGTGATCC CTGAAGCTGT 951
 GCTCAGAGAC CCTTTCCTAG CCGTTCCTGC TCTCTGCTTG CCTCCAAACG CATGCTTCAT 1011
 5 CTGACTTCGG CTTCTCACCT CTGTAGCCTG ACGGACCAAT GCTSCAATGG AAGGGAGGAG 1071
 AGTGATGTGG GGTGCCCCCT CCCTCTCTTC CCTTTGCTTT CCTCTCACTT GGGCCCTTTG 1131
 TGAGATTTTT CTTTGGCCTC CTGTAGAATG GAGCCAGACC ATCCTGGATA ATGTGAGAAC 1191
 ATGCCTAGAT TTACCCACAA AACACAAGTC TGAGAATTAA TCATAAACGG AAGTTTAAAT 1251
 10 GAGGATTTGG ACCTTGGTAA TTGTCCCTGA GTCCTATATA TTTCAACAGT GGCTCTATGG 1311
 GCTCTGATCG AATATCAGTG ATGAAAATAA TAATAATAAT AATAATAACG AATAAGCCAG 1371
 AATCTTGCCA TGAAGCCACA GTGGGGATTC TGGGTTCCAA TCAGAAATGG AGACAAGATA 1431
 15 AAACCTTGCAT ACATTCTTAT GATCACAGAC GGCCCTGGTG GTTTTGGTA ACTATTTACA 1491
 AGGCATTTTT TTACATATAT TTTTGTGCAC TTTTATGTT TCTTTGGAAG ACAAATGTAT 1551
 TTCAGAATAT ATTTGTAGTC AATTCATATA TTTGAAGTGG AGCCATAGTA ATGCCAGTAG 1611
 ATATCICTAT GATCTTGAGC TACTGGCAAC TTGTAAAGAA ATATATATGA CATATAAATG 1671
 20 TATTGTAGCT TTCCGGTGTG AGCCACGGTG TATTTTCCCA CTTGGAATGA AATTGTATCA 1731
 ACTGTGACAT TATATGCACT AGCAATAAAA TGCTAATTGT TTCATGCTGT AAAAAAAAAA 1791
 AAAAAA 1797

Claims

1. A process for constructing a cDNA library comprising:

 - (1) synthesizing single-stranded DNA complementary to mRNA isolated from subject cells using a random polymer and ligating oligo dT to the 3'-end of the single-stranded DNA;
 - (2) converting the single-stranded DNA obtained in (1) to a double-stranded DNA using as primer a poly A oligomer ligated to a first restriction enzyme (enzyme I) site;
 - (3) fragmenting the double-stranded DNA obtained in (2), fractionating the fragments obtained by size, ligating linker containing a second restriction enzyme (enzyme II) site, differing from the enzyme I, thereto and fractionating;
 - (4) amplifying fragments containing the sites of restriction enzymes I and II by a polymerase chain reaction using a first primer containing the enzyme I site and a second primer containing the enzyme II site, digesting the cDNA thus amplified with the restriction enzymes, enzyme I and enzyme II and fractionating; and
 - (5) ligating the cDNA fragment thus obtained upstream of a gene coding for a secretory protein or membrane protein having deleted therefrom the sequence coding for a signal peptide, integrating the ligated DNA into a eucaryotic cell expression plasmid vector, and transforming the vector into a host.
2. A process according to claim 1, wherein EcoRI is used as the enzyme I, SacI is used as the enzyme II, and human IL-2 receptor α gene is used as a known secretory protein or membrane protein gene with the deletion of signal peptide.
3. A polypeptide having the amino acid sequence shown in SEQ ID No. 1 in substantially purified form, a homologue thereof or a fragment of the sequence or homologue of a fragment.
4. A polypeptide according to claim 1 having the amino acid sequence shown in SEQ ID No. 1.
5. A polypeptide comprising a polypeptide according to claim 3 or 4, having the amino acid sequence of residues 1 to 70 shown in SEQ ID No. 4 in substantially purified form, a homologue thereof, or a fragment of the sequence or homologue of a fragment.

- E

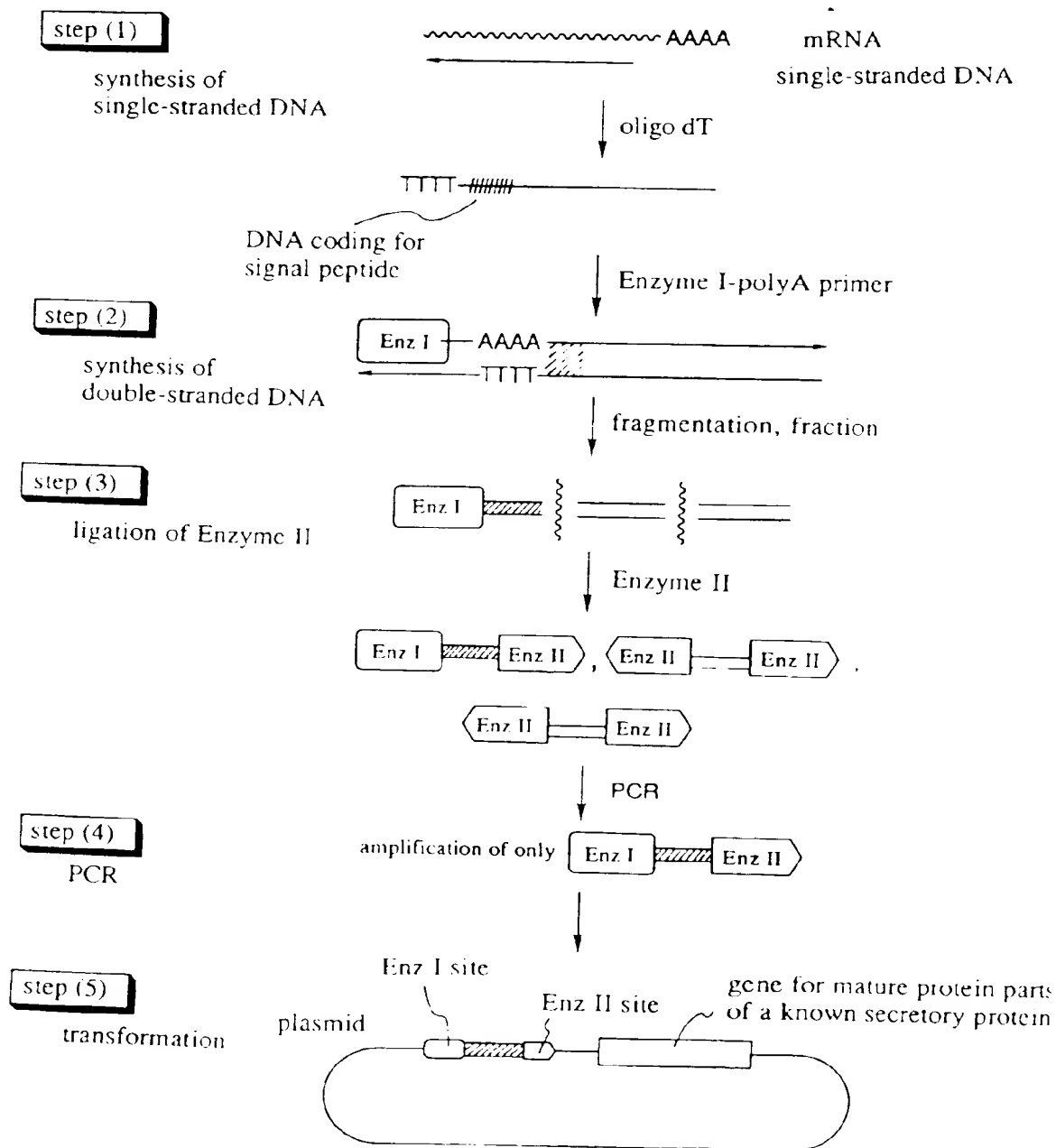
Figure 1

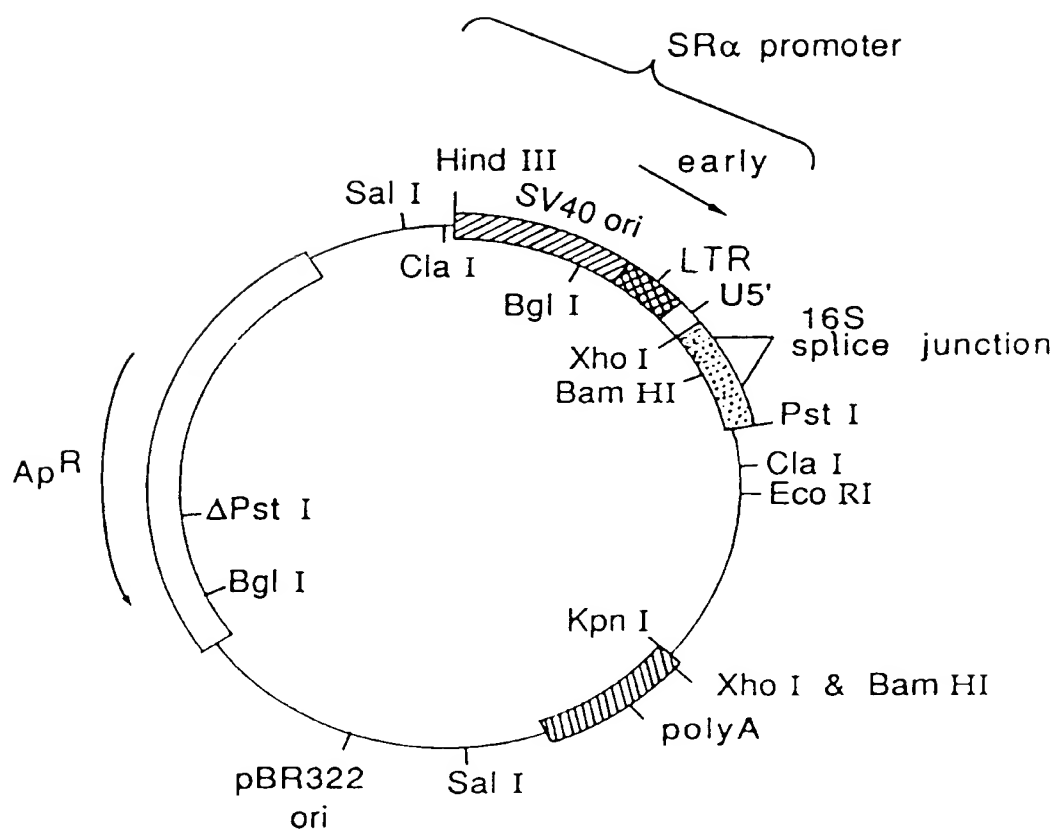
Figure 2

Figure 3

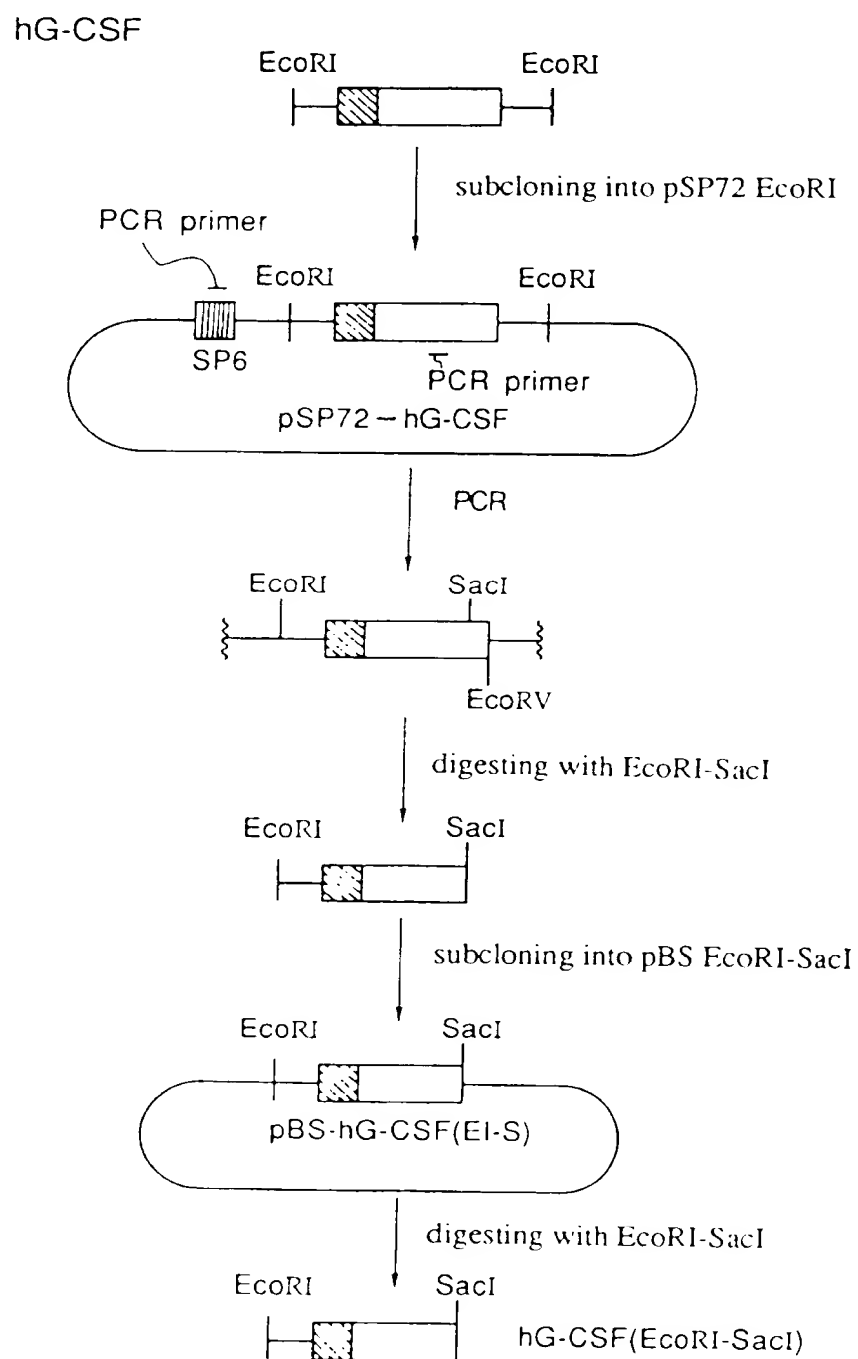


Figure 4

hIL-2R(Tac)

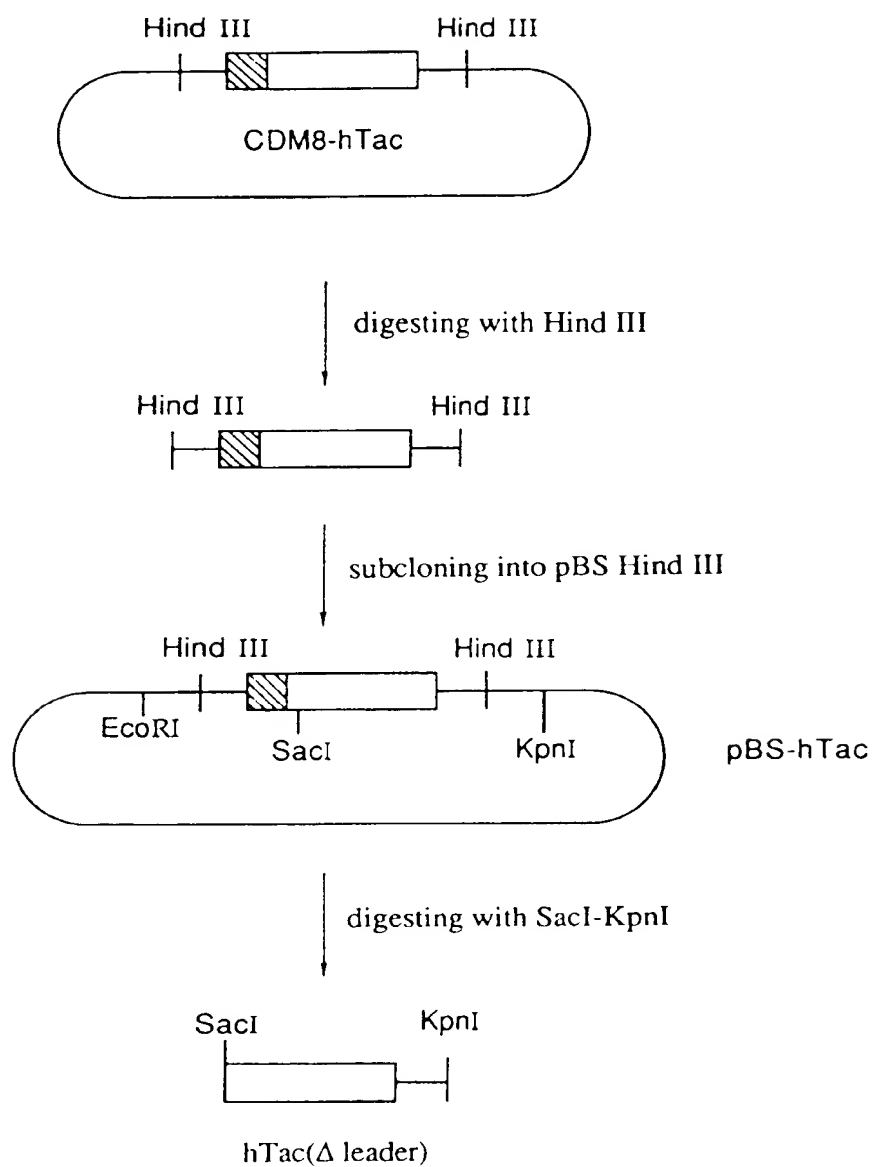
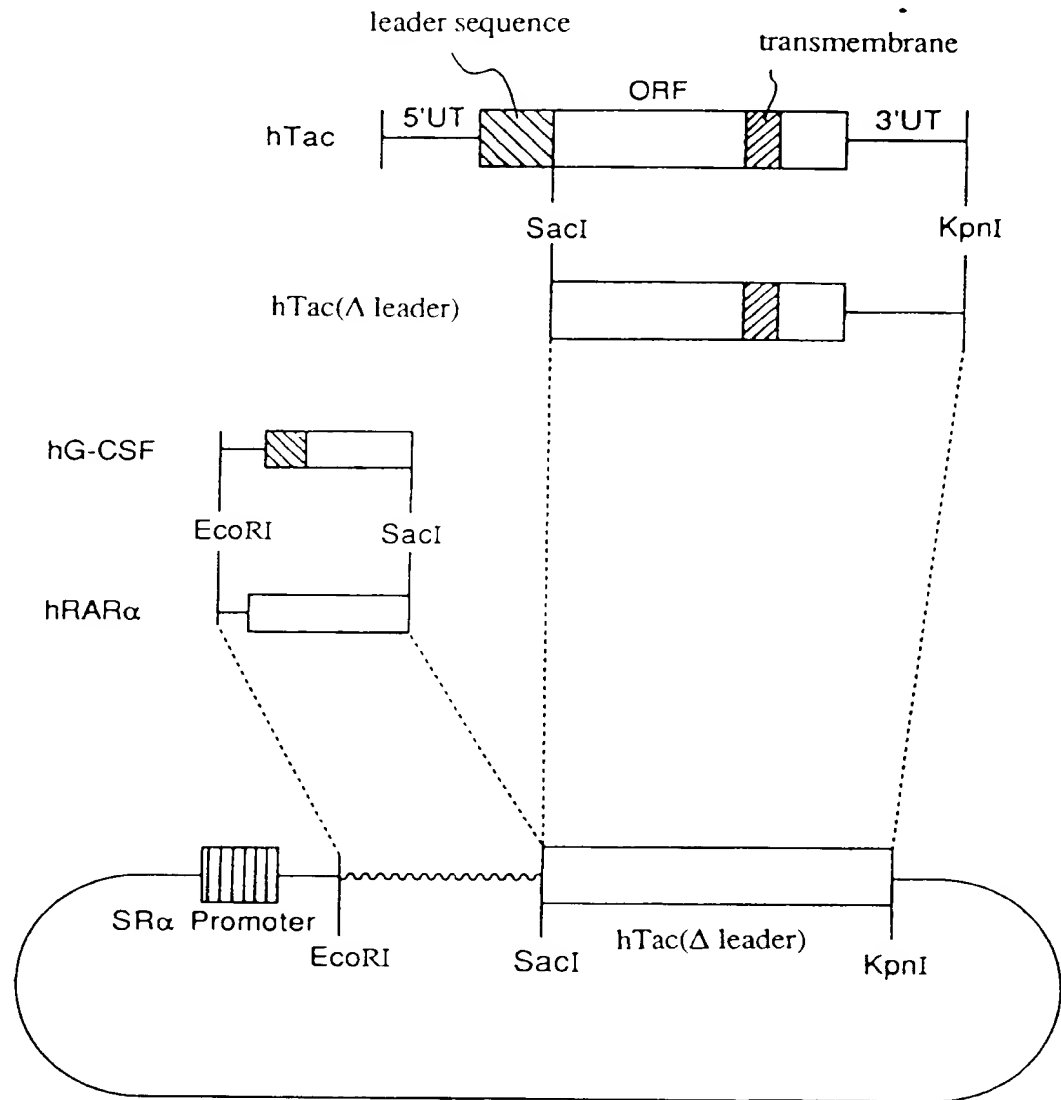


Figure 5



pcDL-SRα-hG-CSF-hTac(Δ leader) (pSGT)

pcDL-SRα-hRARα-hTac(Δ leader) (pSRT)

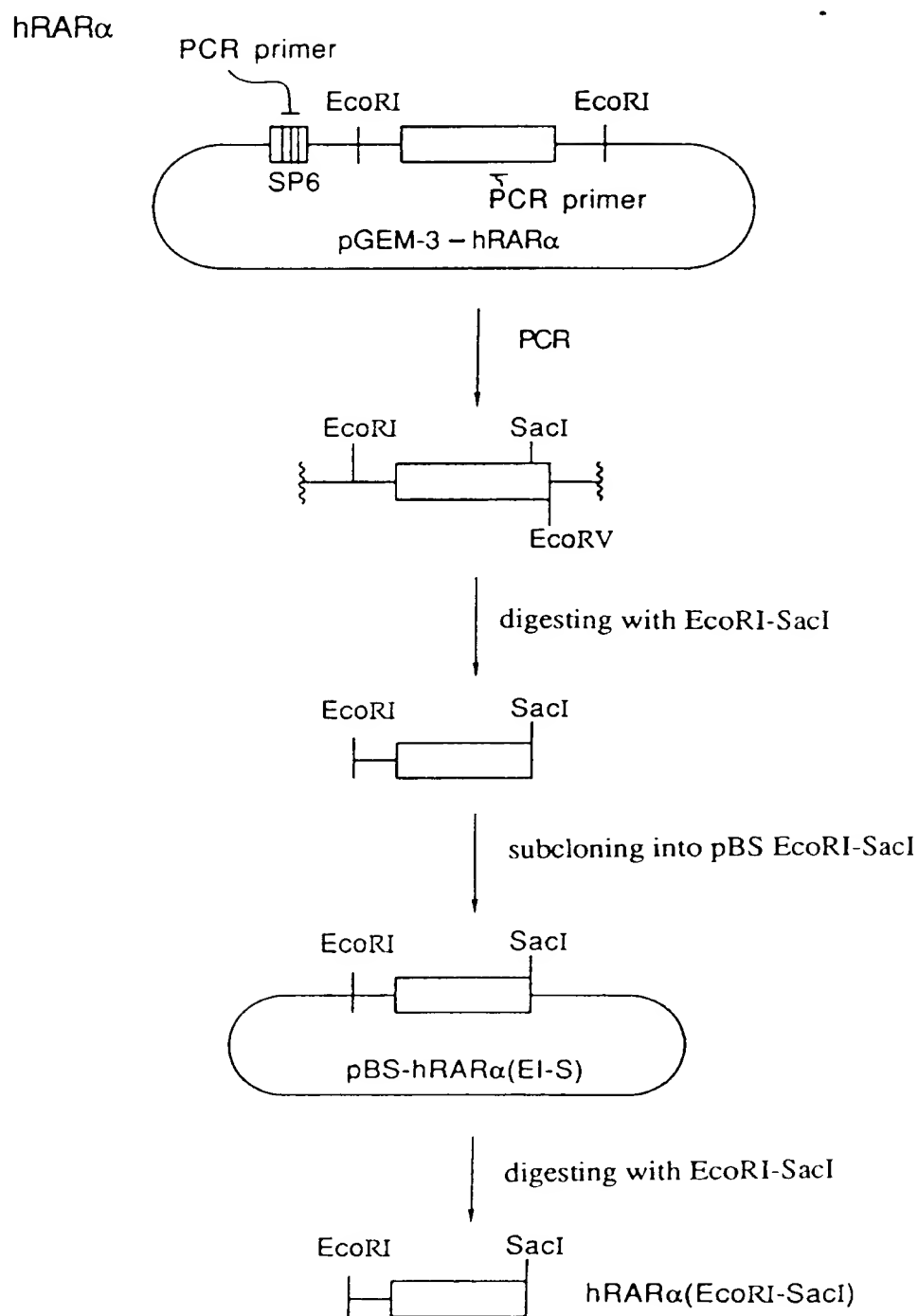
Figure 6

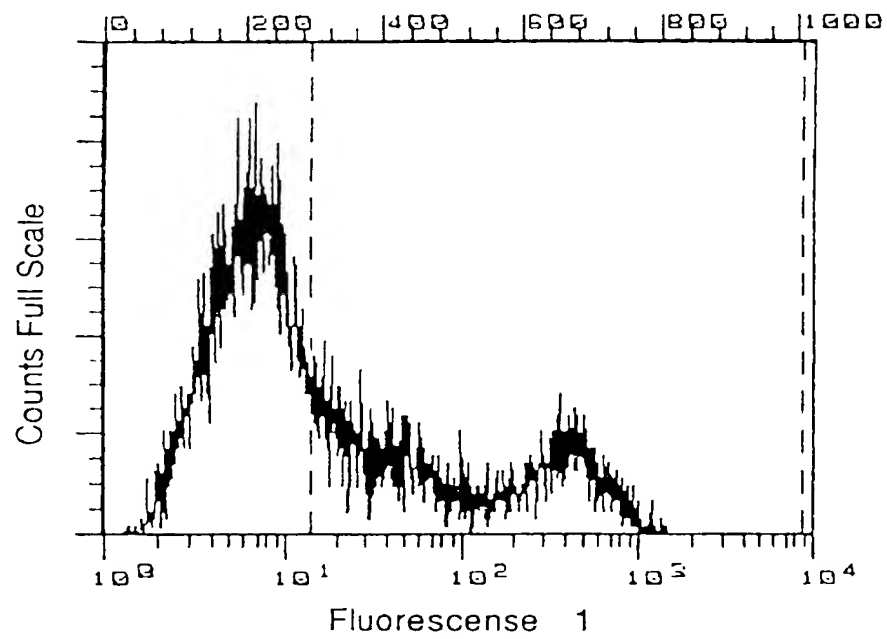
Figure 7

Figure 8

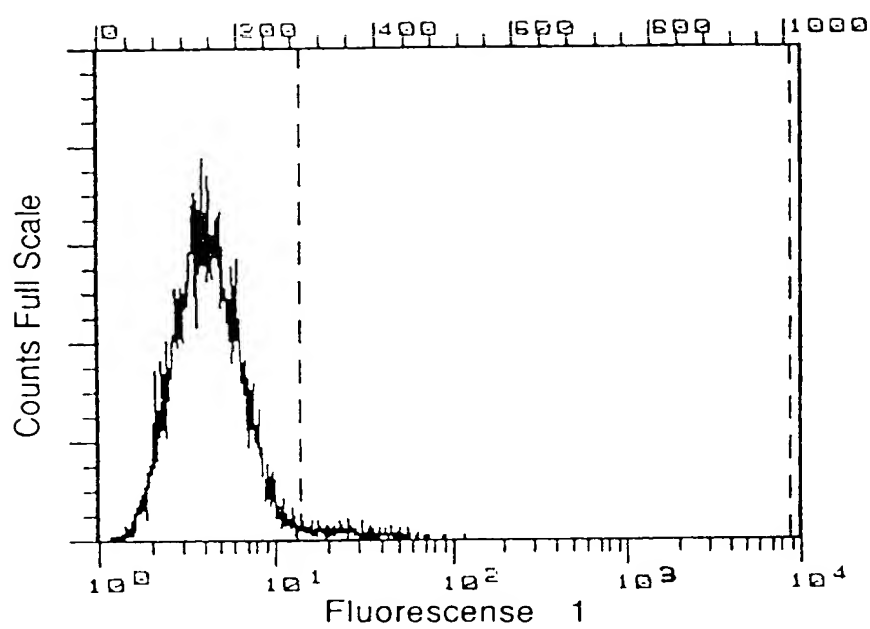


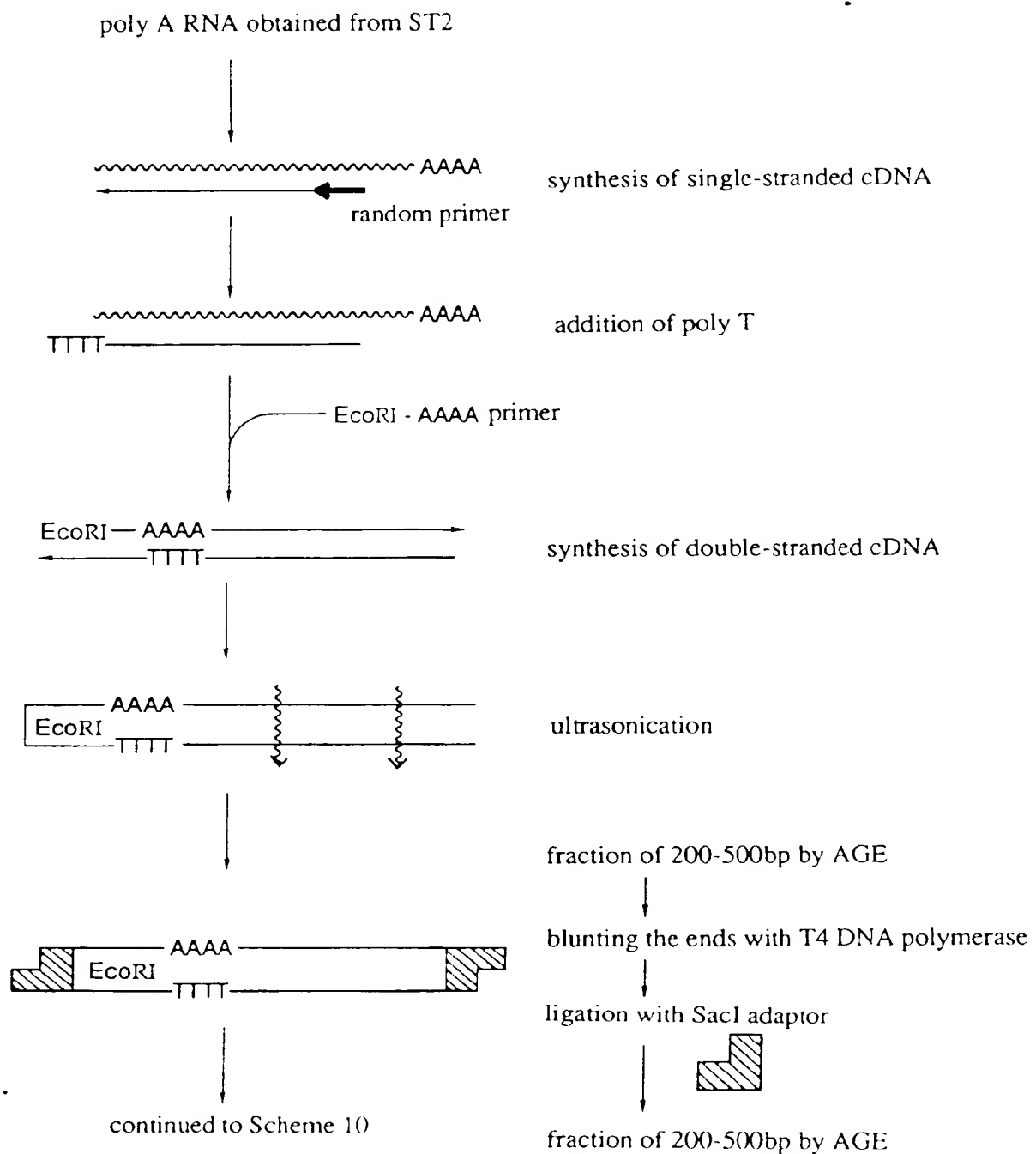
Figure 9

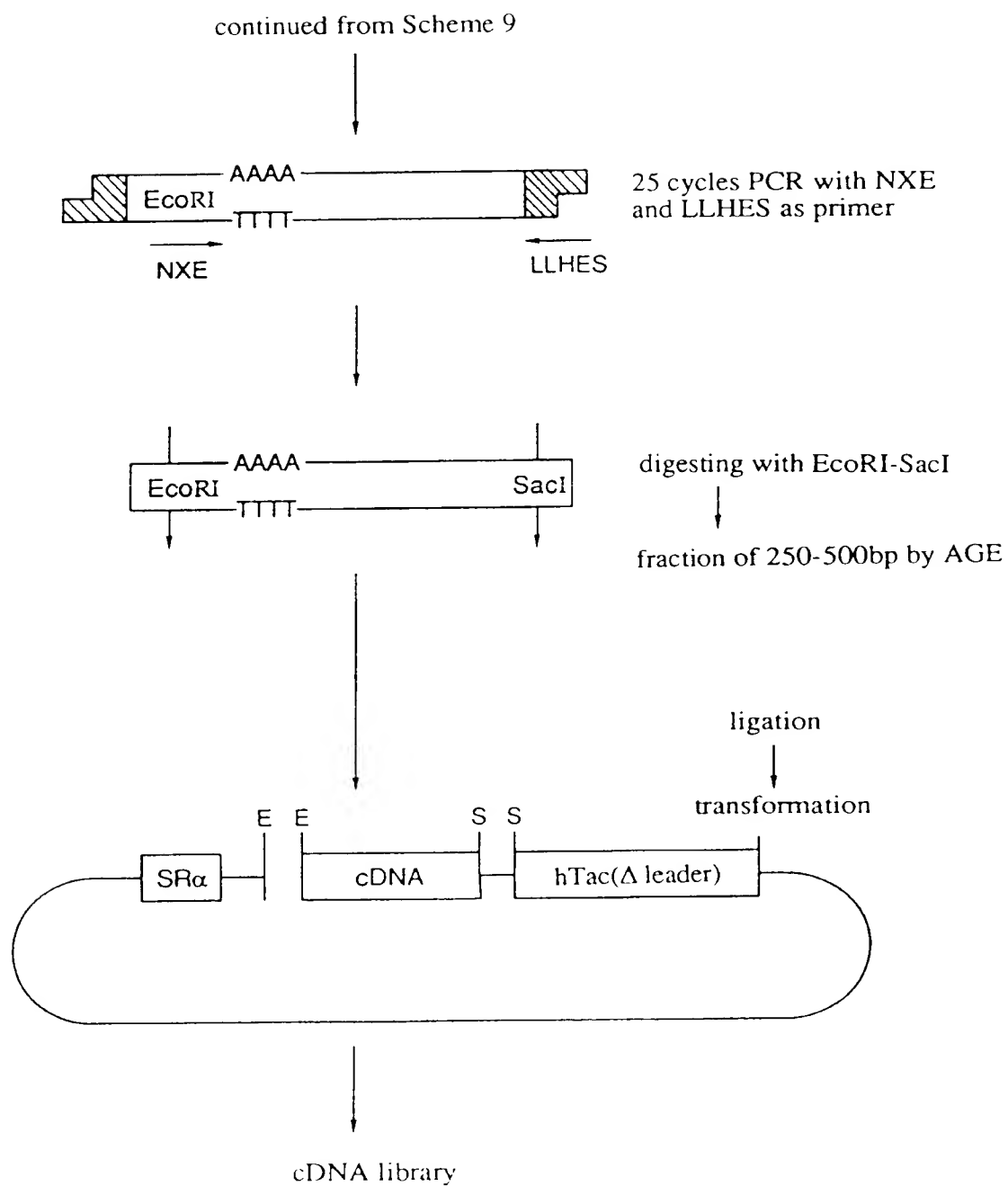
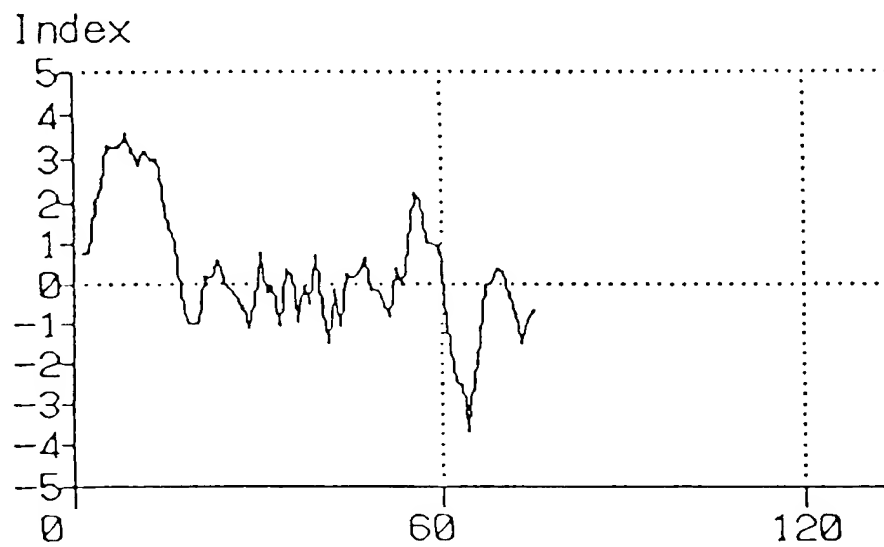
Figure 10

Figure 11



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number: **0 607 054 A3**

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: **94300267.5**

(22) Date of filing: **14.01.94**

(51) Int. Cl.⁵: **C12N 15/10**, C12N 15/12,
C12N 15/62, C12N 15/85,
C07K 13/00, C07K 15/28,
C12Q 1/68, A61K 37/02,
A61K 39/395

(30) Priority: **14.01.93 JP 22098/93**

(43) Date of publication of application:
20.07.94 Bulletin 94/29

(84) Designated Contracting States:
**AT BE CH DE DK ES FR GB GR IE IT LI LU MC
NL PT SE**

(88) Date of deferred publication of search report:
08.02.95 Bulletin 95/06

(71) Applicant: **Honjo, Tasuku**
Kanyu-chi,
Oiwake-cho
Kita-Shirakawa, Sakyo-ku, Kyoto (JP)

(71) Applicant: **ONO PHARMACEUTICAL CO., LTD.**
1-5, Doshomachi 2-chome
Chuo-ku Osaka 541 (JP)

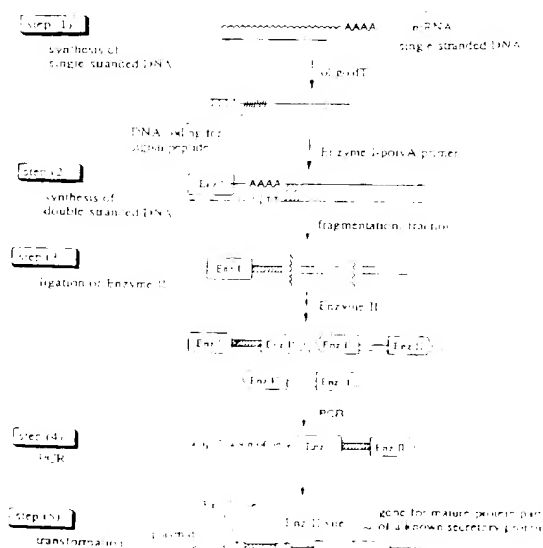
(72) Inventor: **Honjo, Tasuku Kanyu-chi**
Oiwake-cho
Kita-Shirakawa
Sakyo-ku Kyoto (JP)
Inventor: **Tashiro, Kei**
93 Koyama Higashi
Ohno-cho
Kita-ku
Kyoto-shi Kyoto (JP)
Inventor: **Tade, Hideaki c/o Minase Research**
Institute
Ono Pharmaceuticals Co.Ltd.
3-1-1, Sakurai
Shimamoto-cho Mishima-gun Osaka (JP)

(74) Representative: **Bentham, Stephen et al**
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

(54) **Novel process for constructing a cDNA library and a novel polypeptide and DNA coding the same.**

(57) A process for constructing a cDNA library which has a selectivity for signal peptides, that makes it possible to efficiently identify unknown and useful polypeptide comprising a signal peptide. A novel polypeptide consisting of 89 amino acids (including a signal peptide) produced by a stroma cell line, which is useful as an agent for preventing or treating, for example, anemia, leukopenia or infections and the like, and DNAs coding for said polypeptide, have been identified using the process of the invention.

Figure 1



EP 0 607 054 A3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 30 0267

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
Y	NUCL. ACIDS RES., vol.19, no.25, December 1991, IRL PRESS, OXFORD, ENGLAND; pages 7097 - 7104 E. HARA ET AL. 'Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells' * page 7098, left column, line 12 - page 7099, left column, line 15; figure 2 * ---	1	C12N15/10 C12N15/12 C12N15/62 C12N15/85 C07K13/00 C07K15/28 C12Q1/68 A61K37/02 A61K39/395
Y	WO-A-92 18521 (LIFE TECHNOLOGIES, INC.) 29 October 1992 * figures 4,5 * ---	1	
Y	EP-A-0 244 042 (GIST-BROCADES N.V.) 4 November 1987 * the whole document * ---	1	
Y	DE-A-39 01 681 (BEHRINGWERKE AG) 26 July 1990 * claims 6-11 * ---	1	TECHNICAL FIELDS SEARCHED (Int.Cl.5)
P,X	SCIENCE, vol.261, 30 July 1993, AAAS, WASHINGTON, DC, US; pages 600 - 603 K. TASHIRO ET AL. 'Signal sequence trap: A cloning strategy for secreted proteins and type I membrane proteins' * page 600, middle column, line 1 - page 602, right column, line 3 * ---	1-14	C12N C12Q C07K A61K
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 15 December 1994	Examiner Hornig, H
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application I : document cited for other reasons & : member of the same patent family, corresponding document</p>			

EPO FORM 1503 (12-92) (P04/C01)

EUROPEAN SEARCH REPORT

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
T	<p>PROC. NATL. ACAD. SCI., vol.91, no.6, 15 April 1994, NATL. ACAD SCI., WASHINGTON, DC, US; pages 2305 - 2309 T. NAGASAWA ET AL. 'Molecular cloning and structure of a pre-B-cell growth-stimulating factor' * page 2308, right column, line 17 - page 2308, left column, line 48; figure 3 *</p> <p>-----</p>	3-14	
			TECHNICAL FIELDS SEARCHED (Int.Cl.5)
<p>The present search report has been drawn up for all claims</p>			
Place of search THE HAGUE		Date of completion of the search 15 December 1994	Examiner Hornig, H
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p>		<p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>	

